Available online at www.sciencedirect.com
SCIENCE DIRECT
Neural Networks 19 (2006) 113–121
Neural Networks
www.elsevier.com/locate/neunet

2006 Special issue

Computational intelligence in earth sciences and environmental applications: Issues and challenges

V. Cherkassky [a,*], V. Krasnopolsky [b,c], D.P. Solomatine [d], J. Valdes [e]

[a] Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA
[b] SAIC, EMC/NCEP/NOAA, Camp Springs, MD, USA
[c] Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[d] UNESCO-IHE Institute for Water Education, Delft, The Netherlands
[e] National Research Council, Institute for Information Technology, Montreal, Canada

Abstract

This paper introduces a generic theoretical framework for predictive learning, and relates it to data-driven and learning applications in earth and environmental sciences. The issues of data quality, selection of the error function, incorporation of the predictive learning methods into the existing modeling frameworks, expert knowledge, model uncertainty, and other application-domain specific problems are discussed. A brief overview of the papers in the Special Issue is provided, followed by discussion of open issues and directions for future research.

Keywords: Neural networks; Predictive learning; Earth sciences; Environment; Climate; Hydrology

1. Introduction

In this editorial paper, we have attempted to reach the following goals (i) to introduce to practitioners a generic framework of the predictive learning (PL) approach (Section 2); (ii) to introduce a simple classification and a brief review of the PL applications in earth and environmental sciences, and discuss specific issues related to these applications (Section 3); (iii) to briefly overview the papers included in this issue (Section 4); and (iv) to highlight the open issues and future research directions.

2. Framework for predictive learning

The problem of predictive learning (aka inductive learning, machine learning, or learning from examples) can be described in different ways (Mitchell, 1997; Ripley, 1996). In this paper, we adopt the framework of statistical learning (Cherkassky & Mulier, 1998; Friedman, 1994; Vapnik, 1982) shown in Fig. 1.

The setting for predictive learning (PL) involves three components:

- *Generator* of random input vectors $\mathbf{x}$, drawn independently from a fixed (but unknown) probability distribution $P(\mathbf{x})$;
- *System* (or teacher) which returns an output value y for every input vector $\mathbf{x}$ according to the fixed conditional distribution $P(y|\mathbf{x})$, which is also unknown;
- *Learning machine*, which implements a set of approximating functions $f(\mathbf{x},w)$, where w is a set of parameters of an arbitrary nature.

The goal of learning is to select a function (from this set) which approximates best the System's response. This selection is based on the knowledge of finite number (n) of training samples $(\mathbf{x}_i,y_i)$, $(i=1,\ldots,n)$ generated according to (unknown) joint distribution $P(\mathbf{x},y)=P(\mathbf{x})P(y|\mathbf{x})$.

The quality of an approximation produced by the learning machine is measured by the discrepancy or loss $L(y, f(\mathbf{x},\omega))$ between the true output produced by the System and its estimate produced by the learning machine for given input $\mathbf{x}$. By convention, the loss takes on non-negative values, so that large positive values correspond to poor approximation. The expected value of the loss is given by the *prediction risk functional*:

$$R(\omega) = \int L(y, f(\mathbf{x}, \omega))\mathrm{d}P(\mathbf{x}, y) \quad (1)$$

* Corresponding author.
E-mail addresses: cherkass@ece.umn.edu (V. Cherkassky), vladimir.krasnopolsky@noaa.gov (V. Krasnopolsky), d.solomatine@unesco-ihe.org (D.P. Solomatine), julio.valdes@nrc-cnrc.gc.ca (J. Valdes).

0893-6080/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2006.01.001

Learning is the process of finding the function $f(\mathbf{x},\omega_0)$, which minimizes the risk functional (1) over the set of functions supported by the learning machine, using only finite training data (since $P(\mathbf{x},y)$ is unknown). We also point out that the loss function $L(y, f(\mathbf{x},\omega))$ is given a priori based on the problem/application requirements. The prediction risk (1) is unknown, but in practice can be estimated using an independent test set, or via resampling techniques. This formulation (as stated above) is very general and describes many learning problems such as interpolation, regression, classification, and density approximation (Cherkassky & Mulier, 1998; Friedman, 1994; Hastie, Tibshirani, & Friedman, 2001; Vapnik, 1982, 1995).

The problem encountered by the learning machine is to select a function (from the set of functions it supports) that best approximates the System's response. The learning machine is limited to observing finite number (n) examples in order to make this selection. This training data as produced by the generator and system will be independent and identically distributed (iid) according to the joint probability density function (pdf) $p(\mathbf{x},y)$. The finite sample (training data) from this distribution is denoted by:

$$(\mathbf{x}_i, y_i), \quad (i = 1, \ldots, n) \tag{2}$$

With finite data, we cannot expect to find the solution $f(\mathbf{x},\omega_0)$ minimizing prediction risk (1) exactly, so we denote $f(\mathbf{x},\omega^*)$ as the estimate of the optimal solution obtained with finite training data using some learning procedure. It is clear that any learning task (regression, classification, etc.) can be solved by minimizing (1) if the density $p(\mathbf{x},y)$ is known. This means that density estimation is the most general (and hence most difficult) type of learning problem. The problem of learning (estimation) from finite data alone is inherently ill posed. To obtain a useful (unique) solution, the learning process needs to incorporate a priori knowledge in addition to data. For example, such a priori knowledge may be reflected in the set of approximating functions of a learning machine.

Note that a generic learning system shown in Fig. 1 may have two distinct interpretations. Under classical statistical framework, the goal of learning is accurate *identification* of the unknown System, whereas under *predictive learning* (PL) the goal is accurate *imitation* (of a System's output). It should be clear that the goal of system identification is much more demanding than the goal of system imitation. For instance, accurate system identification does not depend on the distribution of input samples; whereas good predictive model is usually conditional upon this (unknown) distribution. Hence,

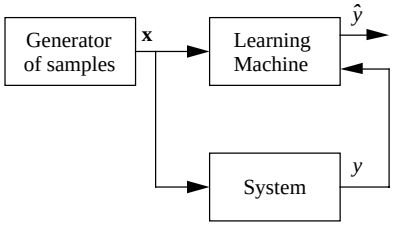

Fig. 1. A learning machine using observations of the system to form an approximation of its output.

an accurate model (in the sense of System's identification) would certainly provide good generalization (in the predictive sense), but the opposite may not be true. The mathematical treatment of system identification leads to the function approximation framework, and to fundamental problems of estimating multivariate functions known as the curse of dimensionality. On the other hand, the goal of accurate system imitation (via minimization of prediction risk) leads to more tractable learning formulations under finite sample settings (Vapnik, 1982, 1995). However, the VC-theoretical approach to PL also requires an appropriate learning problem formulation. This *problem specification* step performs mapping of application-domain requirements onto an appropriate PL formulation, as discussed in Section 3.

Many learning methods are based on the standard (inductive) formulation of the learning problem presented above. For example, a given application is usually formalized as either standard classification or regression problem, even when such standard formulations do not reflect application requirements. Such inductive learning settings assume that:

- the number of future (test) samples is very large, as implied in the expression for risk (1). Moreover, the input ($\mathbf{x}$) values of test samples are unknown during model estimation (training);
- the goal of learning is to model the training data using a single (albeit complex) model;
- the learning machine (in Fig. 1) has a univariate output;
- specific loss functions are used for classification and regression problems.

These assumptions may not hold for many applications. For example, if the input values of the test samples are known (given), then an appropriate goal of learning may be to predict outputs *only* at these points. This leads to the transduction formulation (Vapnik, 1995). Relaxing the assumption about estimating (learning) a single model leads to multiple model estimation formulation (Cherkassky & Ma, 2005). Likewise, it may be possible to relax the assumption about a univariate output under standard supervised learning settings. In many applications, it is necessary to estimate multiple outputs (multivariate functions) of the same input variables. Such methods (for estimating multiple output functions) have been widely used by practitioners, i.e. partial least squares (PLS) regression in chemometrics (Frank & Friedman, 1993). Further, standard loss functions (in classification or regression formulations) may not be appropriate for many applications.

Even though the *problem specification* step cannot be formalized, we suggest several useful guidelines to aid practitioners in the formalization process (Cherkassky, 2001, 2005). The block diagrams for mapping application requirements onto a learning formulation (shown in Fig. 2) advocates the top–down process for specifying three important components of the problem formulation (loss function, input/output variables, and training/test data) based on application needs. In particular, this may include:

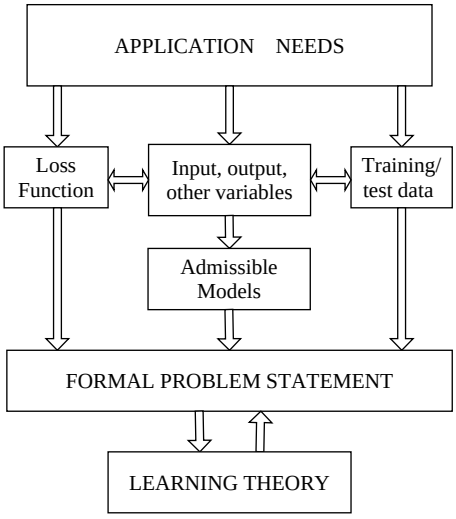

Fig. 2. Mapping application requirements onto a formal learning problem formulation.

- Quantitative or qualitative description of a suitable *loss function*, and relating this loss function to 'standard' learning formulations.
- Description of the *input and output variables*, including their type, range, and other statistical characteristics. In addition to these variables, some applications may have *other variables* that cannot be measured (observed) directly, or can be only partially observed. The knowledge of such variables is also beneficial, as a part of a priori knowledge.
- Detailed characterization of the *training and test data*. Here, the training data refers to the data used for model estimation, and test data denotes future data that will be applied to the model. This includes information about the size of the data sets, knowledge about data generation/collection procedures, etc. More importantly, it is important to describe (and formalize) the use of training and test data in an application-specific context. We emphasize that these notions of training and test data relate to the problem formulation, rather than to a particular learning algorithm. Hence, we do not specify here the *validation* data (a portion of training data used for model complexity control, or tuning hyper-parameters of a learning method). Specification of the validation data set may be a part of the learning algorithm (using data-driven complexity control). However, model complexity control can be performed (in principle) analytically, in which case there is no need for validation data.

Based on a good understanding and specification of these three components, it is usually possible to define a set of admissible models (or approximating functions) shown in Fig. 2. Finally, the formal learning problem statement needs to be related to some theoretical framework (denoted as learning theory in Fig. 2). In practice, this formalization process involves a number of iterations, simplifications and trade-offs. Clear understanding of the components shown in Fig. 2 is useful for understanding the relationship between the learning formulation, application needs, and assumed theoretical paradigm or learning theory. Such an understanding is critical for evaluating the quality of predictive models and interpretation of empirical comparisons between different learning algorithms.

In practice, before data is fed into the learning machine, it has often to be pre-processed, normalized, or undergoes some non-linear transformations (Pyle, 1999). Applications in earth and environmental sciences often have specific characteristics that influence the way the learning problems are posed and solved. An important feature of such problems is that data is either scarce and requires gaps in-filling; or there are so many system variables and/or so many examples that they have to be combined into some aggregates. The following sections will address these and other application-specific issues related to data encoding and pre-processing.

3. Earth sciences and environmental applications: main characteristics

In the 1990s, the field of PL matured; several clear and fundamental textbooks have been published (Beale & Jackson, 1990; Bishop, 1995; Cherkassky & Mulier, 1998; Haykin, 1994; Ripley, 1996; Vapnik, 1995, 1998) that introduced this new powerful statistical (or PL) approach (especially, its particular case—the neural network (NN) technique) that is capable of providing a diverse family of flexible non-linear data-driven models for various applications. The message was sent to the broad community of professionals including scientists working in different fields of geosciences like satellite remote sensing, meteorology, oceanography, geophysical numerical modeling, hydrology, etc. Since then, a significant number of PL applications have been developed in these fields; the most important of them are summarized here:

- Satellite meteorology and oceanography (e.g. classification, pattern recognition, retrieval algorithms, etc.);
- Hybrid climate and weather numerical models and data assimilation systems (e.g. fast emulation of physical processes, fast forward models for data assimilation);
- Geophysical data fusion and data mining;
- Interpolation, downscaling, non-linear multivariate statistical analysis (various areas);
- Hydrologic applications (e.g. modeling rainfall–runoff relationships, flood forecasting, precipitation forecasting).

A number of these applications have been reviewed in several survey papers. Selected atmospheric and oceanic applications have been reviewed by Gardner and Dorling (1998), Krasnopolsky and Chevallier (2003), and Krasnopolsky and Fox-Rabinovitz (2006). Selected remote sensing applications have been reviewed by Atkinson and Tatnall (1997) and Krasnopolsky and Schiller (2003). Applications of the NN technique for developing non-linear generalizations of multivariate statistical analysis have been recently reviewed by

Hsieh (2004). Solomatine (2005) reviewed the applications of machine learning in hydrology.

A large number of important practical applications in environmental and earth sciences (many of applications presented in this issue) may be considered mathematically as a mapping between two vectors **x** (input vector) and **y** (output vector) and can be symbolically written as:

$$\mathbf{y} = M(\mathbf{x}); \quad \mathbf{x} \in \Re^n, \quad \mathbf{y} \in \Re^m \tag{3}$$

For example, a generic application that can be formally considered as a mapping (3) is a retrieval algorithm (or transfer function) in the satellite remote sensing that converts the input vector **x** of satellite measurements (calibrated or raw radiances, brightness temperature, backscatter coefficients, etc. at different frequencies) into the vector **y** of geophysical parameters (wind speeds, atmospheric moisture parameters, ocean and land surface characteristics, etc.). Examples of such applications in this Special Issue include papers by Loyola, and Brajard et al. It is noteworthy that the components of the vector **x** may be intercorrelated because the frequency bands may be not completely independent and overlap. The components of the output vector **y** may be intercorrelated because the corresponding geophysical parameters are physically related. The mapping (3) in this example is usually a complicated non-linear mapping.

Another example of an important application that can be cast as a mapping (3) is a parameterization of atmospheric physics in climate or weather prediction numerical models. Here, the input vector **x** is composed of several functions of height or profile (temperature, humidity, ozone concentration, etc.) and some scalar characteristics. The output vector **y** is composed of functions of height (e.g. the long wave heating rates) and several scalar characteristics. The components of the vector **x** can be significantly intercorrelated and also components of the vector **y** can be significantly correlated because (i) they are physically related and (ii) they are related as the discretized values (elements of profile) of the same continuous function at close values of the argument. The mapping (3) is also a complicated non-linear mapping in this case because the atmospheric physics processes are complicated non-linear ones. The mapping may contain a finite number of finite discontinuities (like step functions) due to intervention of atmospheric moisture processes. This application is presented in this issue by the paper by Krasnopolsky and Fox-Rabinovitz.

Two features mentioned above: (i) correlation between components of input and output vectors and (ii) significant non-linearity of the input/output relationship are quite generic features of environmental and earth science applications.

3.1. Measuring model error

As mentioned in Section 2, the choice of the adequate loss function (model error) is quite important. In modeling natural phenomena, a typical model error function in regression problems is either root mean squared error (RMSE) or similar functions closely related to RMSE, and this fits very well a typical error used in PL. However, there are situations when RMSE, following the principle 'good on average', cannot guarantee that the model performance is optimal in critical situations. For example, in hydrologic forecasting for flood management machine learning models are trained to predict water flow several hours ahead on the basis of past records of rainfall and flow. It is important to ensure that the models perform well during extreme events leading to floods. In this case, the use of RMSE error calculated over an extended modeling period may not be appropriate.

A better alternative in this case is the use of a weighted RMSE where the records corresponding to high rainfall or flow are assigned higher weight. However, during extreme events data usually contain a significantly higher level of observation noise; thus, weighting these data, without clear understanding and taking into account the noise statistics, may lead to an increase in the uncertainty in the NN weights and, therefore, in the uncertainty of NN predictions. An example of training ANN with different loss functions that better reflect the model error in hydrologic context than RMSE is reported, for example, by de De Vos and Rientjes (2005). In this issue, the paper by Dawson, See, Abrahart, and Heppenstall also deals with this issue by introducing a NN optimization routine that is very well fitted for any type of the error function. In classification problems, there is also a need of handling non-standard error functions: the paper by Bhattacharya and Solomatine deals with a particular clustering and classification problem where the standard model error criteria are modified to reflect the contiguity constraints specifically tailored for soil classification applications.

Another alternative is to separate the error calculation for extreme events is training separate models for low and high flows or low and high precipitation conditions, and combining them in a modular scheme; this approach is, for example, reported in this issue by Solomatine and Siek.

3.2. Heterogeneous and complex nature of data

Data characterizing natural phenomena often originates from different heterogeneous sources, ranging from historical paper-based records to fully automated sensors measuring environmental variables. Sometimes the collected data covers long periods (often with gaps, however), but often it is collected during short measurement campaigns. Another problem is that the periods when one group of variables is measured (e.g. water levels or flows) does not necessarily coincide with the periods of other measurements (e.g. temperature or atmospheric pressure). All this leads to the problem of constructing representative training data sets upon which a meaningful predictive model can be estimated.

Other problems include noise in the data (due to unknown sources of errors, as well as measurement errors), impossibility to have full geographical coverage during data collection (in case of building distributed models), and, especially in environmental modeling, lack of relevant data.

There are also applications characterized by very large data sets, especially in oceanography and climate modeling, and in analysis of remote sensing data. In such situations, efficient methods to reduce the problem dimension and to discover the hidden patterns are needed—they would allow for additional analysis and even knowledge discovery. A representative example of such an approach is presented in this issue by Ilin, Valpola, and Oja who apply an advanced version of ICA for data reduction and knowledge discovery.

Databases in geosciences are heterogeneous and incomplete information systems composed of large numbers of objects, described in terms of an increasing number of properties. Locational attributes aside, the properties describing objects are highly *heterogeneous*: some are quantitative, some qualitative, and others more complex, such as time series, spectra, images, etc. Although the quality of information is enriched, the use of such attributes adds complexity and heterogeneity to the analysis. An extra complication in geoscience databases is the problem of *missing data*. In addition, developments in modern monitoring, laboratory and observation equipment lead to increasingly large databases. Moreover, the *uncertainty* in measurements and observations contribute to database complexity.

The development of appropriate machine learning techniques in this case must face the additional problem of accounting for the spatial and temporal dependencies derived from the dynamic nature of earth and environmental processes on one hand, and the geo-referenced relations of the data. The overall consideration of all of these aspects poses considerable challenges.

Because of the sparseness of the environmental data, scarcity of the extreme events, and unavailability of some types of measurements, in many applications data simulated by physically-based models are used for training data-driven (i.e. NN) models. Sometimes, they are used instead of observed data; or they are used to complement the observed data and are integrated with it to form a blended data set. Because the observed and simulated data have different error statistics, the use of a proper blending technique that takes care of different error characteristics is very important. Data assimilation systems extensively used in weather and climate prediction to generate initial conditions for running numerical weather prediction and climate simulation models may serve as an example of such a proper data integration system.

3.3. Fitting machine learning models into existing modeling frameworks

In practice, machine-learning models are to be built in situations when there are already existing models which are typically process models (descriptive, behavioral, and physically-based) and domain experts are trained in using them. Challenge is in introducing very different paradigm of modeling—data-driven modeling based on PL methods—and its incorporation into the existing modeling frameworks.

Traditional complex environmental numerical models are deterministic models based on 'first principle' equations. They are formulated using relevant first principles and observational data, and are usually based on solving deterministic equations (such as radiative transfer equations) and some secondary empirical components based on traditional statistical techniques like regression. Thus, for widely used the state-of-the-art environmental models (like global climate model (GCM) or numerical weather prediction models) all major model components are predominantly deterministic; namely, not only model dynamics but also model physics and chemistry are based on solving deterministic first principle physical or chemical equations.

Only recently attempts have been made to introduce major statistical components into such physically-based models, like an attempt to apply a traditional statistical technique as the expansion of hierarchical correlated functions to approximate atmospheric chemistry components (Schoendorf, Rabitz, & Li, 2003). This traditional statistical technique was applied successfully but had limited accuracy. Significantly, higher accuracy requirements must be met for such complex multidimensional and interdisciplinary systems as modern environmental and earth sciences models. A particular type of PL technique, namely NN technique, has been successfully applied for the development of new and for emulation of existing atmospheric and ocean physics parameterizations (Chevallier, Chéruy, Scott, & Chédin, 1998; Krasnopolsky, Breaker, & Gemmill, 1997; Krasnopolsky, Chalikov, & Tolman, 2002; Krasnopolsky, Fox-Rabinovitz, & Chalikov, 2005). Hybrid numerical models were introduced which are based on a synergetic combination of deterministic numerical modeling with PL modules emulating model components (Krasnopolsky & Fox-Rabinovitz, 2006).

3.4. Emulation of physically-based process models

Sometimes the learning models are used not to emulate a modeled phenomenon directly but to mimic or emulate a process (physically-based) model of this phenomenon, that is to perform meta-modeling. Such models are also called surrogate models, and they are trained on the data generated or simulated by process models.

In hydrology and hydraulics, two applications can be mentioned. Solomatine and Torres (1996) used hydrologic and hydrodynamic models of river flows to generate enough data to train a neural network that would forecast the water levels. This fast-running replica was used during solving a problem of optimizing the water reservoir operation. Another example is reported by Khu, Savic, Liu, and Madsen (2002) where a NN replica of a hydrologic model was used to accelerate the process of its calibration.

Such an approach has also been developed in modeling atmospheric and oceanic processes (Chevallier et al., 1998; Krasnopolsky et al., 1997, 2000; Krasnopolsky et al., 2005) (presented in this journal issue by Krasnopolsky and Fox-Rabinovitz). Following this approach, authors developed NN emulations that are from one to five orders of magnitude faster

than original physically based models. The authors also showed that combining these PL (NN) components with deterministic, physically-based ones and creating fast hybrid numerical modes offers new opportunities for numerical climate and weather prediction.

Another example is reported in the paper by Loyola in this issue where an ensemble of neural networks is used to emulate complex radiative transfer models for the purpose of modeling ozone column behavior. The resulting learning model runs much faster than the simulation model, which is very important for real-time ozone monitoring using satellites.

Note, that as explained in Section 2, such surrogate models do not attempt to perform the system identification, but only to approximate (imitate) the system output.

3.5. Challenges in acceptance of data-driven models

Due to the characteristics of the earth and environmental applications, domain experts typically constitute an important part of the whole modeling cycle. A challenge is to increase the role of domain experts in tuning the learning models. Some PL algorithms directly involve experts in the process of building models: for building decision trees (Ankerst et al., 1999), or for model trees (Solomatine & Siek, this issue).

Criticism often heard from the builders and users of process models is that PL models are 'black-boxes' that do not reflect the physics of the modeled process. In other words, the PL models are imitation models rather than system identification models (see Section 2). Indeed, equations induced by PL models have very different form if compared to the equations describing the physics (or chemistry) of the modeled processes. Clear understanding and analysis of the underlying physical processes is very important PL modeling. The results of such analysis are reflected in the relevant input and output variables (see Fig. 2).

In hydrological applications, for example, good understanding of the modeled hydrological unit (catchment) helps in the choice of the input variables. If the problem is river flow forecasting, increase in flow now is induced by the increased rainfall in the past, and the good knowledge of the catchment should be the basis for choosing the proper lags with which rainfall time series are brought into the PL model. Additionally, correlation analysis and mutual information can be used for variable selection.

There is no doubt that the black-box nature of the data-driven models provided by most machine learning procedures is one of the most important factors conditioning the reluctance of domain experts to accept them, even when they exhibit good performance. For example, understanding neural network based models may be obscured by the intricacies of its architecture and the sheer number of its parameters (weights). In the case of a fuzzy system, the set of fuzzy rules might be large and complex. Moreover, the number of linguistic variables required and the collection of membership functions might be large as well. In this respect, attention should be paid to the use of meta-learning techniques aiming at (i) analyzing black-box models with the purpose of deriving explanatory knowledge about their properties and operation, and (ii) direct construction of deterministic models described by analytical functions.

Analytic functions describing with physical systems in general, have had a long history in science. They are easier to understand by humans, the preferred building blocks of modeling, and a highly condensed form of knowledge. Regression is an example where the family of functions is restricted to a few ones (typically just one), and the problem reduces to finding a set of parameters or coefficients, which make the function, fulfill some desirable approximation property (for example, minimizing a least squares error). A possible approach could be training a number of simple (linear) functions, comprising an overall non-linear mapping (application of this approach in hydrology has been reported by Solomatine and Dulal (2003), and is also addressed in this issue by Solomatine and Siek). However, direct discovery of general analytic functions poses enormous challenges because of, among many other problems, the (potentially) infinite size of the search space, and the need of approximate arbitrary complex non-linearities.

In this respect, the developments in evolutionary computation algorithms, in particular those within the branch of genetic programming, have an interesting potential (see, e.g. Liong et al., 2002). The analytical expressions obtained may approximate a broad scope of non-linearities, and the analysis of the impact of the input variables in the model can be directly examined by looking at the nature of the functional dependencies in which the different variables are involved. Moreover, sensitivity analysis is considerably simplified because of the existence of an explicit functional relationship. However, often the analytic expressions found are very complex, and since for a given problem there might be many different functions satisfying the same approximation constraints, the functions found may mask simpler functional links between the input and the target variables, which would be more meaningful to a domain expert.

3.6. Dealing with model uncertainty

The PL (NN) technique is a non-linear statistical approach. As any statistical approach, the NN technique is expected to provide not only a minimization of an error function and a single-point prediction, but also an estimate of the uncertainties in the model weights and outputs (predictions). Because of the non-linear nature of the NN technique, an estimation of the NN uncertainties is a more complicated business than that in the case of linear statistical tools and models. However, during last decade a progress has been made in this filed both for the case of NNs with a single output (Bishop, 1996; Heskes, 1997; Nabney, 2002; Neal, 1996; Nix & Weigend, 1994) and with multiple outputs (Aires, Prigent, & Rossow, 2004). Various Bayesian methods have been used in these studies to estimate the uncertainties for the NN parameters (weights). In this issue, a new method to estimate the prediction uncertainty is proposed by Shrestha and Solomatine.

It is a fact a life that regardless of the generalization performance or interpretability of a prediction model, from the point of view of a decision maker, the value of a prediction depends on the availability of additional information regarding the dependability of the prediction, and risks associated with decisions taken upon this prediction. In earth and environmental sciences, such risks associated with decisions based on model predictions are especially large. Therefore, deriving appropriate techniques for estimating uncertainties associated with predictions provided by data-driven models is of utmost importance.

4. Papers in this issue

The contributions to this issue are clustered according to three application-domains as follows.

4.1. Climate

This section includes four contributions. The paper by Vladimir Krasnopolsky, Michael Fox-Rabinovitz and Ming-Dah Chou deals with the use of NNs in emulating very complex models atmospheric long wave radiation: their approach considerably reduced the total running time of a global climate model. The paper by Yonas Dibike and Paulin Coulibaly successfully applied temporal NN to the important problem of downscaling the global climate models outputs in order to use them in the climate change impact studies. The paper by Aiming Wu and William Hsieh deals with the forecasts of the tropical pacific sea surface temperature based on an ensemble of NNs. All three of the mentioned papers demonstrate high effectiveness of applying different types of NNs (or their ensembles as in the paper by Wu and Hsu) to either predict environmental variables, or to emulate a process model. The paper by Alexander Ilin, Harri Valpola, and Erkki Oja describes the application of a new powerful extension of PCA (denoising source separation) to the climate data analysis. Their approach shows the possibility of discovering patterns in climate data, in this case leading to a very interesting result of detecting a well-known effect of El Nino—being the component with the highest interannual variability.

4.2. Earth and ocean

Since the global processes on the Earth and in the ocean directly influence the climate, this section closely relates to the previous one. It includes five contributions. The paper by Diego Loyola reviews various types of model mixtures and applies modular NNs to the processing of earth observation satellite data—for the determination of cloud properties and for the retrieval of total columns of ozone. The developed algorithms are currently being used for the operational processing of European atmospheric satellite sensors. The contribution by Julien Brajard, Cedric Jamet, Cyril Moulin, and Sylvie Thiria addresses an issue of improving the estimates of the chlorophyll concentrations in the ocean based on the satellite on-board sensors measuring the solar radiation reflected by the ocean and the atmosphere. Such estimates are based on the models of radiative transfer simulation, and NNs were successfully used as such models. The paper by Biswa Bhattacharya and Dimitri Solomatine deals with the issue of training NN and SVM classification models that would predict the soil types on the basis of so-called cone penetration tests that generate data on the soil friction and resistance. The paper by Julio Valdes and Graeme Bonham-Carter addresses an issue of detecting the state changes in complex processes (with application to the paleoclimate and solar data) by generating collections of time-dependent non-linear autoregressive models represented by a special kind of neuro-fuzzy NNs that are generated by model-mining procedures. The paper by Biswa Bhattacharya and Dimitri Solomatine deals with a problem of predicting the sedimentation in a large port 3 weeks in advance. The predictive models were NNs and M5 model trees and one of the problems to solve was the identification of the proper time lags for the variables characterizing wind, waves and river flow and data pre-processing based on the inclusion of the physical processes of sedimentation.

4.3. Hydrology

The three contributions in this section deal typically with the processes having a shorter time scale. In the paper by Dimitri Solomatine and Michael Siek, a number of approaches to combining modular learning models are presented (including the new algorithms for optimization of building M5 piece-wise liner model trees), and their use is illustrated by building the models for the water flow predictions. The paper by Durga Lal Shrestha and Dimitri Solomatine addresses the issue of model uncertainty by introducing and testing a new method for estimating the prediction intervals for the model outputs; its usefulness is demonstrated on the rainfall–runoff NN models. The paper by Christian Dawson, Linda See, Bob Abrahart, and Alison Heppenstall describes a symbiotic adaptive neuro-evolution method advancing on traditional evolutionary approaches by evolving and optimizing individual neurons. The method is applied to rainfall–runoff modeling, with the use of alternative model error functions that better fit the hydrological situation. This section of the issue reflects some of the latest developments in and applications of predictive learning to water-related issues, that are often associated with the area of hydroinformatics. The authors focus on relatively non-standard ways of building predictive models for solving the water-related issues—local modeling of input sub-spaces and fuzzy combination of the resulting models, adaptive evolutionary optimization of NNs, and constructing the predictive models of model uncertainty.

5. Conclusion and open issues

Applications of new modeling methods in relation to the earth and environmental sciences have high social, humanitarian and economic value. Predictive learning provides a powerful framework for building effective data-driven models

that can complement and in many instances even replace the traditional process models. Often these models are faster than the process numerical models and can be used as their replicas that are included into the process modeling frameworks leading to the hybrid models, or included in the model-based optimization procedures.

There are a number of specific characteristics of such applications, as detailed in Section 3. Even though predictive learning methods often yield very effective models, their use is never straightforward, and one should pay special attention to incorporating prior knowledge (about underlying physical processes) into the learning problem formulation, selection of appropriate learning methods, and evaluation of modeling results. For example, standard (RMSE) error functions used in the NN 'textbook' training procedures sometimes have to be updated to reflect better typical (heavy-tail) error distributions, or instead of a single model a number of local models have to be trained that would better reflect the heterogeneous nature of the data.

There are a number of open issues related to the applications of PL in earth and environmental sciences. In our view, the most important ones are the following:

1. The quality of data in these applications is often lower than in many other applications (for example, those in industry). So the challenge is to choose/design robust learning methods that can handle better heterogeneous data, missing data and non-standard noise.
2. Acceptance of PL methods in the communities traditionally accustomed to process models. An important practical issue is the incorporation of PL methods into existing modeling frameworks, including better incorporation of application–domain knowledge. In many applications in earth and environmental sciences, predictive models are actually used not for prediction per se, but for policy/decision-making (i.e. debate on global warming). So the notions such as the model interpretability and model uncertainty become very important. These issues, strictly speaking, lie outside the scope of traditional PL framework, which is concerned only with prediction accuracy (generalization). Hence, more research is definitely needed in this direction.
3. In earth and environmental sciences, there is a growing demand for having predictions that have some sort of uncertainty estimates associated with them. The approaches based on the Monte-Carlo sampling and running the process models could be computationally too demanding to be practical. PL methods make it possible to build models of uncertainty trained on the historical data and the database of the (process) model runs.

In conclusion, we emphasize that earth and environmental sciences represent an important practical domain for machine learning and predictive learning methods. This SI provides a systematic view of this field, and hopefully, will lead to further advances and concentrated research efforts along the issues outlined above.

References

Aires, F., Prigent, C., & Rossow, W. B. (2004). Neural network uncertainty assessment using Bayesian statistics: A remote sensing application. *Neural Computation*, *16*, 2415–2458.

Atkinson, P. M., & Tatnall, A. R. L. (1997). Neural networks in remote sensing—Introduction. *International Journal of Remote Sensing*, *18*(4), 699–709.

Beale, R., & Jackson, T. (1990). *Neural computing: An introduction* (240 pp). Bristol: Adam Hilger.

Bishop, C. M. (1995). *Neural networks for pattern recognition* (482 pp). Oxford, UK: Oxford University Press.

Cherkassky, V. (2001). New formulations for predictive learning, plenary lecture *ICANN 2001*, Vienna, Austria.

Cherkassky, V. (2005). *New Formulations for Predictive Learning, IJCNN-05 tutorial*, available at http://ebrains.la.asu.edu/~jennie/tutorial/tutorial.htm.

Cherkassky, V, & Ma, Y (2005). Multiple model regression estimation. *IEEE Transactions on Neural Networks*, *16*(4), 785–798.

Cherkassky, V., & Mulier, F. (1998). *Learning from data: Concepts, theory and methods*. New York: Wiley.

Chevallier, F., Chéruy, F., Scott, N. A., & Chédin, A. (1998). A neural network approach for a fast and accurate computation of longwave radiative budget. *Journal of Applied Meteorology*, *37*, 1385–1397.

De Vos, N. J., & Rientjes, T. H. M. (2005). Constraints of artificial neural networks for rainfall–runoff modelling: Trade-offs in hydrological state representation and model evaluation. *Hydrology and Earth System Sciences*, *9*, 111–126.

Frank, I., & Friedman, J. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, *35*(2), 109–148 (with discussion).

Friedman, J. (1994). An overview of predictive learning and function approximation. In V. Cherkassky, J. Friedman, & H. Wechsler (Eds.), *From statistics to neural networks. NATO ASI series F: Vol. 136*. New York: Springer.

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*, 2627–2636.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference and prediction*. New York: Springer.

Haykin, S. (1994). *Neural networks: A comprehensive foundation* (pp. 696). New York: Macmillan.

Hsieh, W. W. (2004). Nonlinear multivariate and time series analysis by neural network methods. *Reviews of Geophysics*, *42*, RG1003.

Khu, S. T., Savic, D., Liu, Y. & Madsen, H. (2002). A fast evolutionary-based meta-modelling approach for the calibration of a rainfall–runoff model. In: Pahl-Wostl, Schmidt, Rizzoli, & Jakeman, (Eds.). *Proceedings of 1st biennial meeting of the international environmental modelling and software society* (Vol. 1) (pp. 147–152), iEMSs.

Krasnopolsky, V., Breaker, L.C., & Gemmill, W.H. (1997). *A neural network forward model for direct assimilation of SSM/I brightness temperatures into atmospheric models. Research activities in atmospheric and oceanic modeling, CAS/JSC Working Group on Numerical Experimentation, Report no. 25, WMO/TD-No. 792* (pp. 129–130).

Krasnopolsky, V. M., Chalikov, D. V., & Tolman, H. L. (2002). A neural network technique to improve computational efficiency of numerical oceanic models. *Ocean Modelling*, *4*, 363–383.

Krasnopolsky, V. M., & Chevallier, F. (2003). Some neural network applications in environmental sciences. Part II: Advancing computational efficiency of environmental numerical models. *Neural Networks*, *16*, 335–348.

Krasnopolsky, V. M., & Fox-Rabinovitz, M. S. (2006). A new synergetic paradigm in environmental numerical modeling: Hybrid models combining deterministic and machine learning components. *Ecological Modelling*, *191*, 5–18.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Chalikov, D. V. (2005). New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of long wave radiation in a climate model. *Monthly Weather Review*, *133*, 1370–1383.

Krasnopolsky, V. M., & Schiller, H. (2003). Some neural network applications in environmental sciences. Part I: Forward and inverse problems in satellite remote sensing. *Neural Networks*, *16*, 321–334.

Liong, S. Y., Gautam, T. R., Khu, S. T., Babovic, V., Keijzer, M., & Muttil, N. (2002). Genetic programming: A new paradigm in rainfall runoff modeling. *Journal of the American Water Resources Association*, *38*(3), 705–718.

Mitchell, T. M. (1997). *Machine learning*. New York: McGraw-Hill.

Nabney, I. T. (2002). *Netlab: Algorithms for pattern recognition*. New York: Springer.

Neal, R. M. (1996). *Bayesian learning for neural networks*. New York: Springer.

Pyle, D. (1999). *Data preparation for data mining*. San Francisco, CA: Morgan Kaufmann.

Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Schoendorf, J., Rabitz, H., & Li, G. (2003). A fast and accurate operational model of ionospheric electron density. *Geophysical Research Letters*, *30*, 1492–1495.

Solomatine, D. P. (2005). Data-driven modeling and computational intelligence methods in hydrology. In M. Anderson (Ed.), *Encyclopedia of hydrological sciences*. New York: Wiley.

Solomatine, D. P., & Dulal, K. N. (2003). Model tree as an alternative to neural network in rainfall–runoff modelling. *Hydrological Sciences Journal*, *48*(3), 399–411.

Solomatine, D. P., & Torres, L. A. (1996). Neural network approximation of a hydrodynamic model in optimizing reservoir operation. *Proceeding of the second international conference on hydroinformatics* (pp. 201–206). Rotterdam: Balkema.

Vapnik, V. (1982). *Estimation of dependences based on empirical data*. Berlin: Springer.

Vapnik, V. (1995). *The nature of statistical learning theory*. Berlin: Springer.

Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.